

# Decision Making under Interval Uncertainty (and Beyond)

Vladik Kreinovich  
Department of Computer Science  
University of Texas at El Paso  
500 W. University  
El Paso, Texas 79968, USA  
vladik@utep.edu

## Abstract

To make a decision, we must find out the user's preference, and help the user select an alternative which is the best – according to these preferences. Traditional utility-based decision theory is based on a simplifying assumption that for each two alternatives, a user can always meaningfully decide which of them is preferable. In reality, often, when the alternatives are close, the user is often unable to select one of these alternatives. In this chapter, we show how we can extend the utility-based decision theory to such realistic (interval) cases.

## 1 Introduction

To make a decision, we must:

- find out the user's preference, and
- help the user select an alternative which is the best – according to these preferences.

Traditional utility-based decision theory is based on a simplifying assumption that for each two alternatives  $A'$  and  $A''$ , a user can always meaningfully decide which of them is preferable. In reality, often, when the alternatives are close, the user is often unable to select one of these alternatives. How can we extend the utility-based decision theory to such realistic cases?

In this chapter, we provide an overview of such an extension. This paper is structured as follows: first, we recall the main ideas and results of the traditional utility-based decision theory. We then consider the case when in addition to deciding which of the two alternatives is better, the user can also reply that he/she is unable to decide between the two close alternatives; this leads to interval uncertainty.

*Comment.* Some of the results presented in this paper were previously reported at conferences [1, 22].

## 2 Traditional utility-based decision theory: brief reminder

Following [8, 26, 34], let us describe the main ideas and results of the traditional decision theory.

### Main assumption behind the traditional utility-based decision theory.

Let us assume that for every two alternatives  $A'$  and  $A''$ , a user can tell:

- whether the first alternative is better for him/her; we will denote this by  $A'' < A'$ ;
- or the second alternative is better; we will denote this by  $A' < A''$ ;
- or the two given alternatives are of equal value to the user; we will denote this by  $A' = A''$ .

*Comment.* In mathematical terms, we assume that the preference relation  $<$  is a linear (total) order; in economics, this property of the preference relation is also known as *completeness*.

**The notion of utility.** Under the above assumption, we can form a natural numerical scale for describing attractiveness of different alternatives. Namely, let us select a very bad alternative  $A_0$  and a very good alternative  $A_1$ , so that most other alternatives are better than  $A_0$  but worse than  $A_1$ .

Since we assumed that the alternatives between which we need to choose are linearly ordered, there exists the best one – which can be selected as  $A_1$ , and the worst one – which can be selected as  $A_0$ . However, since one of the main objectives of this paper is to go beyond this simplifying linearity assumption, it is better to select  $A_1$  and  $A_0$  beyond the available alternative. For example, we can choose, as  $A_1$ , an alternative “I win a billion dollars” – we do not have this alternative in our decision, but this alternative is easy to imagine. Similarly, as  $A_0$ , we can select a really bad alternative – and it is OK if this alternative is not a possible outcome of our current decision-making process.

Then, for every probability  $p \in [0, 1]$ , we can form a lottery  $L(p)$  in which we get  $A_1$  with probability  $p$  and  $A_0$  with the remaining probability  $1 - p$ .

When  $p = 0$ , this lottery simply coincides with the alternative  $A_0$ :  $L(0) = A_0$ . The larger the probability  $p$  of the positive outcome increases, the better the result, i.e.,  $p' < p''$  implies  $L(p') < L(p'')$ . Finally, for  $p = 1$ , the lottery coincides with the alternative  $A_1$ :  $L(1) = A_1$ . Thus, we have a continuous scale of alternatives  $L(p)$  that monotonically goes from  $A_0$  to  $A_1$ .

We have assumed that most alternatives  $A$  are better than  $A_0$  but worse than  $A_1$ :  $A_0 < A < A_1$ . Since  $A_0 = L(0)$  and  $A_1 = L(1)$ , for such alternatives, we thus get  $L(0) < A < L(1)$ . We assumed that every two alternatives can be compared. Thus, for each such alternative  $A$ , there can be at most one value  $p$  for which  $L(p) = A$ ; for others, we have  $L(p) < A$  or  $L(p) > A$ . Due to monotonicity of  $L(p)$  and transitivity of preference, if  $L(p) < A$ , then  $L(p') < A$  for all  $p' \leq p$ ; similarly, if  $A < L(p)$ , then  $A < L(p')$  for all  $p' > p$ . Thus, the supremum (= least upper bound)  $u(A)$  of the set of all  $p$  for which  $L(p) < A$  coincides with the infimum (= greatest lower bound) of the set of all  $p$  for which  $A < L(p)$ . For  $p < u(A)$ , we have  $L(p) < A$ , and for  $p > u(A)$ , we have  $A < L(p)$ . This value  $u(A)$  is called the *utility* of the alternative  $A$ .

It may be possible that  $A$  is equivalent to  $L(u(A))$ ; however, it is also possible that  $A \neq L(u(A))$ . However, the difference between  $A$  and  $L(u(A))$  is extremely small: indeed, no matter how small the value  $\varepsilon > 0$ , we have  $L(u(A) - \varepsilon) < A < L(u(A) + \varepsilon)$ . We will describe such (almost) equivalence by  $\equiv$ , i.e., we write that  $A \equiv L(u(A))$ .

**How can we actually find utility values.** The above definition of utility is somewhat theoretical, but in reality, utility can be found reasonably fast by the following iterative bisection procedure.

We want to find the probability  $u(A)$  for which  $L(u(A)) \equiv A$ . On each stage of this procedure, we have the values  $\underline{u} < \bar{u}$  for which  $L(\underline{u}) < A < L(\bar{u})$ . In the beginning, we have  $\underline{u} = 0$  and  $\bar{u} = 1$ , with  $|\bar{u} - \underline{u}| = 1$ .

To find the desired probability  $u(A)$ , we compute the midpoint  $\tilde{u} = \frac{\underline{u} + \bar{u}}{2}$  and compare the alternative  $A$  with the corresponding lottery  $L(\tilde{u})$ . Based on our assumption, there are three possible results of this comparison:

- if the user concludes that  $L(\tilde{u}) < A$ , then we can replace the previous lower bound  $\underline{u}$  with the new one  $\tilde{u}$ ;
- if the user concludes that  $A < L(\tilde{u})$ , then we can replace the original upper bound  $\bar{u}$  with the new one  $\tilde{u}$ ;
- finally, if  $A = L(\tilde{u})$ , this means that we have found the desired probability  $u(A)$ .

In this third case, we have found  $u(A)$ , so the procedure stops. In the first two cases, the new distance between the bounds  $\underline{u}$  and  $\bar{u}$  is the half of the original distance. By applying this procedure  $k$  times, we get values  $\underline{u}$  and  $\bar{u}$  for which  $L(\underline{u}) < A < L(\bar{u})$  and  $|\bar{u} - \underline{u}| \leq 2^{-k}$ . One can easily check that the desired value  $u(A)$  is within the interval  $[\underline{u}, \bar{u}]$ , so the midpoint  $\tilde{u}$  of this interval is an  $2^{-(k+1)}$ -approximation to the desired utility value  $u(A)$ .

In other words, for any given accuracy, we can efficiently find the corresponding approximation to the utility  $u(A)$  of the alternative  $A$ .

**How to make a decision based on utility values.** If we know the utilities  $u(A')$  and  $u(A'')$  of the alternatives  $A'$  and  $A''$ , then which of these alternatives should we choose?

By definition of utility, we have  $A' \equiv L(u(A'))$  and  $A'' \equiv L(u(A''))$ . Since  $L(p') < L(p'')$  if and only if  $p' < p''$ , we can thus conclude that  $A'$  is preferable to  $A''$  if and only if  $u(A') > u(A'')$ .

In other words, we should always select an alternative with the largest possible value of utility.

*Comment.* Interval techniques can help in finding the optimizing decision; see, e.g., [27].

**How to estimate utility of an action: why expected utility.** To apply the above idea to decision making, we need to be able to compute utility of different actions. For each action, we usually know possible outcomes  $S_1, \dots, S_n$ , and we can often estimate the probabilities  $p_1, \dots, p_n$ ,  $\sum_{i=1}^n p_i = 1$ , of these outcomes. Let  $u(S_1), \dots, u(S_n)$  be utilities of the situations  $S_1, \dots, S_n$ . What is then the utility of the action?

By definition of utility, each situation  $S_i$  is equivalent (in the sense of the relation  $\equiv$ ) to a lottery  $L(u(S_i))$  in which we get  $A_1$  with probability  $u(S_i)$  and  $A_0$  with the remaining probability  $1 - u(S_i)$ . Thus, the action in which we get  $S_i$  with probability  $p_i$  is equivalent to complex lottery in which:

- first, we select one of the situations  $S_i$  with probability  $p_i$ :  $P(S_i) = p_i$ ;
- then, depending on the selected situation  $S_i$ , we get  $A_1$  with probability  $u(S_i)$  and  $A_0$  with probability  $1 - u(S_i)$ :  $P(A_1 | S_i) = u(S_i)$  and  $P(A_0 | S_i) = 1 - u(S_i)$ .

In this complex lottery, we end up either with the alternative  $A_1$  or with the alternative  $A_0$ . The probability of getting  $A_1$  can be computed by using the complete probability formula:

$$P(A_1) = \sum_{i=1}^n P(A_1 | S_i) \cdot P(S_i) = \sum_{i=1}^n u(S_i) \cdot p_i.$$

Thus, the original action is equivalent to a lottery in which we get  $A_1$  with probability  $\sum_{i=1}^n p_i \cdot u(S_i)$  and  $A_0$  with the remaining probability. By definition of utility, this means that the utility of our action is equal to  $\sum_{i=1}^n p_i \cdot u(S_i)$ .

In probability theory, this sum is known as the expected value of utility  $u(S_i)$ . Thus, we can conclude that the utility of each action is equal to its expected utility; in other words, among several possible actions, we should select the one with the largest value of expected utility.

**Non-uniqueness of utility.** The above definition of utility depends on a selection of two alternatives  $A_0$  and  $A_1$ . What if we select different alternatives  $A'_0$  and  $A'_1$ ? How will utility change? In other words, if  $A$  is an alternative with utility  $u(A)$  in the scale determined by  $A_0$  and  $A_1$ , what is its utility  $u'(A)$  in the scale determined by  $A'_0$  and  $A'_1$ ?

Let us first consider the case when  $A'_0 < A_0 < A_1 < A'_1$ . In this case, since  $A_0$  is in between  $A'_0$  and  $A'_1$ , for each of them, there exists a probability  $u'(A_0)$  for which  $A_0$  is equivalent to a lottery  $L'(u'(A_0))$  in which we get  $A'_1$  with probability  $u'(A_0)$  and  $A'_0$  with the remaining probability  $1 - u'(A_0)$ . Similarly, there exists a probability  $u'(A_1)$  for which  $A_1$  is equivalent to a lottery  $L'(u'(A_1))$  in which we get  $A'_1$  with probability  $u'(A_1)$  and  $A'_0$  with the remaining probability  $1 - u'(A_1)$ .

By definition of the utility  $u(A)$ , the original alternative  $A$  is equivalent to a lottery in which we get  $A_1$  with probability  $u(A)$  and  $A_0$  with the remaining probability  $1 - u(A)$ . Here,  $A_1$  is equivalent to the lottery  $L'(u'(A_1))$ , and  $A_0$  is equivalent to the lottery  $L'(u'(A_0))$ . Thus, the alternative  $A$  is equivalent to a complex lottery, in which:

- first, we select  $A_1$  with probability  $u(A)$  and  $A_0$  with probability  $1 - u(A)$ ;
- then, depending on the selection  $A_i$ , we get  $A'_1$  with probability  $u'(A_i)$  and  $A'_0$  with the remaining probability  $1 - u'(A_i)$ .

In this complex lottery, we end up either with the alternative  $A'_1$  or with the alternative  $A'_0$ . The probability  $u'(A) = P(A'_1)$  of getting  $A'_1$  can be computed by using the complete probability formula:

$$\begin{aligned} u'(A) &= P(A'_1) = P(A'_1 | A_1) \cdot P(A_1) + P(A'_1 | A_0) \cdot P(A_0) = \\ &= u'(A_1) \cdot u(A) + u'(A_0) \cdot (1 - u(A)) = \\ &= u(A) \cdot (u'(A_1) - u'(A_0)) + u'(A_0). \end{aligned}$$

Thus, the original alternative  $A$  is equivalent to a lottery in which we get  $A'_1$  with probability  $u'(A) = u(A) \cdot (u'(A_1) - u'(A_0)) + u'(A_0)$ . By definition of utility, this means that the utility  $u'(A)$  of the alternative  $A$  in the scale determined by the alternatives  $A'_0$  and  $A'_1$  is equal to  $u'(A) = u(A) \cdot (u'(A_1) - u'(A_0)) + u'(A_0)$ .

Thus, in the case when  $A'_0 < A_0 < A_1 < A'_1$ , when we change the alternatives  $A_0$  and  $A_1$ , the new utility values are obtained from the old ones by a linear transformation. In other cases, we can use auxiliary events  $A''_0$  and  $A''_1$  for which  $A''_0 < A_0, A'_0$  and  $A_1, A'_1 < A''_1$ . In this case, as we have proven, transformation from  $u(A)$  to  $u''(A)$  is linear and transformation from  $u'(A)$  to  $u''(A)$  is also linear. Thus, by combining linear transformations  $u(A) \rightarrow u''(A)$  and  $u''(A) \rightarrow u'(A)$ , we can conclude that the transformation  $u(A) \rightarrow u'(A)$  is also linear.

So, in general, utility is defined modulo an (increasing) linear transformation  $u' = a \cdot u + b$ , with  $a > 0$ .

*Comment.* So far, once we have selected alternatives  $A_0$  and  $A_1$ , we have defined the corresponding utility values  $u(A)$  only for alternatives  $A$  for which  $A_0 < A < A_1$ . For such alternatives, the utility value is always a number from the interval  $[0, 1]$ .

For other alternatives, we can define their utility  $u'(A)$  with respect to different pairs  $A'_0$  and  $A'_1$ , and then apply the corresponding linear transformation to re-scale to the original units. The resulting utility value  $u(A)$  can now be an arbitrary real number.

**Subjective probabilities.** In our derivation of expected utility, we assumed that we know the probabilities  $p_i$  of different outcomes. In practice, we often do not know these probabilities, we have to rely on a subjective evaluation of these probabilities. For each event  $E$ , a natural way to estimate its subjective probability is to compare the lottery  $\ell(E)$  in which we get a fixed prize (e.g., \$1) if the event  $E$  occurs and 0 if it does not occur, with a lottery  $\ell(p)$  in which we get the same amount with probability  $p$ . Here, similarly to the utility case, we get a value  $ps(E)$  for which  $\ell(E)$  is (almost) equivalent to  $\ell(ps(E))$  in the sense that  $\ell(ps(E) - \varepsilon) < \ell(E) < \ell(ps(E) + \varepsilon)$  for every  $\varepsilon > 0$ . This value  $ps(E)$  is called the *subjective probability* of the event  $E$ ; see, e.g., [5, 24, 26, 36].

For each event  $E$ , we can efficiently find its subjective probability by using a bisection procedure which is similar to how we can find utilities.

From the viewpoint of decision making, each event  $E$  is equivalent to an event occurring with the probability  $ps(E)$ . Thus, if an action has  $n$  possible outcomes  $S_1, \dots, S_n$ , in which  $S_i$  happens if the event  $E_i$  occurs, then the utility of this action is equal to  $\sum_{i=1}^n ps(E_i) \cdot u(S_i)$ .

### 3 Towards a more realistic way to describe user preference: interval uncertainty

**Beyond traditional utility-based decision making: towards a more realistic description.** Previously, we assumed that a user can always decide which of the two alternatives  $A'$  and  $A''$  is better:

- either  $A' < A''$ ,
- or  $A'' < A'$ ,
- or  $A' \equiv A''$ .

In practice, a user is sometimes unable to meaningfully decide between the two alternatives  $A'$  and  $A''$ ; see, e.g., [9, 26]. We will denote this option by  $A' \parallel A''$ .

In mathematical terms, this means that the preference relation is no longer a *total* (linear) order, it can be a *partial* order.

**From utility to interval-valued utility.** Similarly to the traditional utility-based decision making approach, we can select two alternatives  $A_0 < A_1$  and compare each alternative  $A$  which is better than  $A_0$  and worse than  $A_1$  with lotteries  $L(p)$ . The main difference is that here, the supremum  $\underline{u}(A)$  of all the values  $p$  for which  $L(p) < A$  is, in general, smaller than the infimum  $\bar{u}(A)$  of all the values  $p$  for which  $A < L(p)$ . Thus, for each alternative  $A$ , instead of a single value  $u(A)$  of the utility, we now have an *interval*  $[\underline{u}(A), \bar{u}(A)]$  such that:

- if  $p < \underline{u}(A)$ , then  $L(p) < A$ ;
- if  $p > \bar{u}(A)$ , then  $A < L(p)$ ; and
- if  $\underline{u}(A) < p < \bar{u}(A)$ , then  $A \parallel L(p)$ .

We will call this interval the *utility* of the alternative  $A$ .

**How to efficiently elicit the interval-valued utility from the user.** To elicit the corresponding utility interval from the user, we can use a slightly modified version of the above bisection procedure. At first, the procedure is the same as before: namely, we produce a narrowing interval  $[\underline{u}, \bar{u}]$  for which  $L(\underline{u}) < A < L(\bar{u})$ .

We start with the interval  $[\underline{u}, \bar{u}] = [0, 1]$ , and we repeatedly compute the midpoint  $\tilde{u} = \frac{\underline{u} + \bar{u}}{2}$  and compare  $A$  with  $L(\tilde{u})$ . If  $L(\tilde{u}) < A$ , we replace  $\underline{u}$  with  $\tilde{u}$ ; if  $A < L(\tilde{u})$ , we replace  $\bar{u}$  with  $\tilde{u}$ . If we get  $A \parallel L(\tilde{u})$ , then we switch to the new second stage of the iterative algorithm. Namely, now, we have *two* intervals:

- an interval  $[\underline{u}_1, \bar{u}_1]$  (which is currently equal to  $[\underline{u}, \tilde{u}]$ ) for which  $L(\underline{u}_1) < A$  and  $L(\bar{u}_1) \parallel A$ , and
- an interval  $[\underline{u}_2, \bar{u}_2]$  (which is currently equal to  $[\tilde{u}, \bar{u}]$ ) for which  $L(\underline{u}_2) \parallel A$  and  $A < L(\bar{u}_2)$ .

Then, we perform bisection of each of these two intervals. For the first interval, we compute the midpoint  $\tilde{u}_1 = \frac{\underline{u}_1 + \bar{u}_1}{2}$ , and compare the alternative  $A$  with the lottery  $L(\tilde{u}_1)$ :

- if  $L(\tilde{u}_1) < A$ , then we replace  $\underline{u}_1$  with  $\tilde{u}_1$ ;
- if  $L(\tilde{u}_1) \parallel A$ , then we replace  $\bar{u}_1$  with  $\tilde{u}_1$ .

As a result, after  $k$  iterations, we get the value  $\underline{u}(A)$  with accuracy  $2^{-k}$ .

Similarly, for the second interval, we compute the midpoint  $\tilde{u}_2 = \frac{\underline{u}_2 + \bar{u}_2}{2}$ , and compare the alternative  $A$  with the lottery  $L(\tilde{u}_2)$ :

- if  $L(\tilde{u}_2) \parallel A$ , then we replace  $\underline{u}_2$  with  $\tilde{u}_2$ ;
- if  $A < L(\tilde{u}_2)$ , then we replace  $\bar{u}_2$  with  $\tilde{u}_2$ .

As a result, after  $k$  iterations, we get the value  $\bar{u}(A)$  with accuracy  $2^{-k}$ .

*Comment.* Similar to the case of exactly known utilities, when we replace alternatives  $A_0$  and  $A_1$  with alternatives  $A'_0$  and  $A'_1$ , the new values  $\underline{u}'$  and  $\bar{u}'$  are related to the original values  $\underline{u}$  and  $\bar{u}$  by the same linear transformation  $u' = a \cdot u + b$ :  $\underline{u}' = a \cdot \underline{u} + b$  and  $\bar{u}' = a \cdot \bar{u} + b$ .

**Interval-valued subjective probability.** Similarly, when we are trying to estimate the probability of an event  $E$ , we no longer get a single value  $ps(E)$ , we get an interval  $[\underline{ps}(E), \bar{ps}(E)]$  of possible values of probability.

By using bisection, we can feasibly elicit the values  $\underline{ps}(E)$  and  $\bar{ps}(E)$ ; alternative ways of eliciting interval-valued probabilities are described in [13, 14].

## 4 Decision making under interval uncertainty

**Need for decision making under interval uncertainty.** In the traditional utility-based approach, for each alternative  $A$ , we produce a number  $u(A)$  – the utility of this alternative. Then, an alternative  $A'$  is preferable to the alternative  $A''$  if and only if  $u(A') > u(A'')$ .

How can we make a similar decision in situations when we only know interval-valued utilities?

*Comment.* Several approaches have been proposed for such decision-making; for example, several approaches for decision making under interval-valued probabilities are described and compared in [41]. In this chapter, we concentrate on approaches which naturally extend the above utility approach.

**How to make a decision under interval uncertainty: a natural idea.**

For each possible decision  $d$ , we know the interval  $[\underline{u}(d), \bar{u}(d)]$  of possible values of utility. Which decision shall we select? A seemingly natural idea is to select all decisions  $d_0$  that *may* be optimal, i.e., which are optimal for some function  $u(d) \in [\underline{u}(d), \bar{u}(d)]$ . There is a minor problem with this definition: that checking all possible functions is not feasible. However, this problem is easy to solve, since this condition can be reformulated in simpler equivalent terms.

Let us describe this reformulation.

**Definition 1.** Let  $D$  be a set; its elements will be called possible decisions. Let  $\mathbf{u}$  be a function that assigns, to each possible decision  $d \in D$ , an interval  $\mathbf{u}(d) = [\underline{u}(d), \bar{u}(d)]$ . A function  $u$  which maps  $D$  into real numbers is called a possible utility function if  $\underline{u}(d) \leq u(d) \leq \bar{u}(d)$  for all  $d$ . We say that a decision  $d_0$  is possibly optimal if  $u(d_0) = \max_{d \in D} u(d)$  for some possible utility function  $u$ .

**Proposition.** A decision  $d_0$  is possibly optimal if and only if

$$\bar{u}(d_0) \geq \max_d \underline{u}(d).$$



*Comment.* This equivalent inequality is indeed easy to check.

**Proof.** If  $d_0$  is possibly optimal, then  $u(d_0) \geq u(d)$  for all  $d$ . Thus, from  $\bar{u}(d_0) \geq u(d_0) \geq u(d) \geq \underline{u}(d)$ , we conclude that  $\bar{u}(d_0) \geq \underline{u}(d)$  for all  $d$ . Hence, we get  $\bar{u}(d_0) \geq \max_d \underline{u}(d)$ .

Vice versa, suppose that  $\bar{u}(d_0) \geq \max_d \underline{u}(d)$ , i.e., that  $\bar{u}(d_0) \geq \underline{u}(d)$  for all  $d$ . Then, we can take the following possible utility function  $u$ :  $u(d_0) = \bar{u}(d_0)$  and  $u(d) = \underline{u}(d)$  for all  $d \neq d_0$ . For this possible utility function,  $u(d_0) \geq u(d)$  for all  $d$ , so  $d_0$  is indeed a possibly optimal decision. The equivalence is proven.

*Comment.* Interval computations can help in describing the range of all such  $d_0$ ; see, e.g., [27].

**Need for definite decision making.** In practice, we would like to select *one* decision; which one should be select?

At first glance, the situation may sound straightforward: if  $A' \parallel A''$ , it does not matter whether we select  $A'$  or  $A''$ . However, this is *not* a good way to make a decision. For example, let us assume that there is an alternative  $A$  about which we know nothing. In this case, we have no reason to prefer  $A$  or  $L(p)$ , so we have  $A \parallel L(p)$  for all  $p$ . By definition of  $\underline{u}(A)$  and  $\bar{u}(A)$ , this means that we have  $\underline{u}(A) = 0$  and  $\bar{u}(A) = 1$ , i.e., the alternative  $A$  is characterized by the utility interval  $[0, 1]$ .

In this case, the alternative  $A$  is indistinguishable both from a good lottery  $L(0.999)$  (in which the good alternative  $A_1$  appears with probability 99.9%) and from a bad lottery  $L(0.001)$  (in which the bad alternative  $A_0$  appears with probability 99.9%). If we recommend, to the user, that  $A$  is equivalent both to  $L(0.999)$  and  $L(0.001)$ , then this user will feel comfortable exchanging his chance to play in the good lottery with  $A$ , and then – following the same logic – exchanging  $A$  with a chance to play in a bad lottery. As a result, following our recommendations, the user switches from a very good alternative to a very bad one.

This argument does not depend on the fact that we assumed complete ignorance about  $A$ . Every time we recommend that the alternative  $A$  is equivalent to  $L(p)$  and  $L(p')$  with two different values  $p < p'$ , we make the user vulnerable to a similar switch from a better alternative  $L(p')$  to a worse one  $L(p)$ . Thus, there should be only a single value  $p$  for which  $A$  can be reasonably exchanged with  $L(p)$ .

In precise terms: we start with the utility interval  $[\underline{u}(A), \bar{u}(A)]$ , and we need to select a single utility value  $u$  for which it is reasonable to exchange the alternative  $A$  with a lottery  $L(u)$ . How can we find this value  $u$ ?

**How to make decisions under interval uncertainty: Hurwicz optimism-pessimism criterion.** The problem of decision making under such interval uncertainty was first handled by the future Nobelist L. Hurwicz in [15].

We need to assign, to each interval  $[\underline{u}, \bar{u}]$ , a utility value  $u(\underline{u}, \bar{u})$ .

No matter what value  $u$  we get from this interval, this value will be larger than or equal to  $\underline{u}$  and smaller than or equal to  $\bar{u}$ . Thus, the equivalent utility value  $u(\underline{u}, \bar{u})$  must satisfy the same inequalities:  $\underline{u} \leq u(\underline{u}, \bar{u}) \leq \bar{u}$ . In particular, for  $\underline{u} = 0$  and  $\bar{u} = 1$ , we get  $0 \leq \alpha_H \leq 1$ , where we denoted  $\alpha_H \stackrel{\text{def}}{=} u(0, 1)$ .

We have mentioned that the utility is determined modulo a linear transformation  $u' = a \cdot u + b$ . It is therefore reasonable to require that the equivalent utility does not depend on what scale we use, i.e., that for every  $a > 0$  and  $b$ , we have

$$u(a \cdot \underline{u} + b, a \cdot \bar{u} + b) = a \cdot u(\underline{u}, \bar{u}) + b.$$

In particular, for  $\underline{u} = 0$  and  $\bar{u} = 1$ , we get

$$u(b, a + b) = a \cdot u(0, 1) + b = a \cdot \alpha_H + b.$$

So, for every  $\underline{u}$  and  $\bar{u}$ , we can take  $b = \underline{u}$ ,  $a = \bar{u} - \underline{u}$ , and get

$$u(\underline{u}, \bar{u}) = \underline{u} + \alpha_H \cdot (\bar{u} - \underline{u}) = \alpha_H \cdot \bar{u} + (1 - \alpha_H) \cdot \underline{u}.$$

This expression is called *Hurwicz optimism-pessimism criterion*, because:

- when  $\alpha_H = 1$ , we make a decision based on the most optimistic possible values  $u = \bar{u}$ ;
- when  $\alpha_H = 0$ , we make a decision based on the most pessimistic possible values  $u = \underline{u}$ ;
- for intermediate values  $\alpha_H \in (0, 1)$ , we take a weighted average of the optimistic and pessimistic values.

So, if we have two alternatives  $A'$  and  $A''$  with interval-valued utilities  $[\underline{u}(A'), \bar{u}(A')]$  and  $[\underline{u}(A''), \bar{u}(A'')]$ , we recommend an alternative for which the equivalent utility value is the largest. In other words, we recommend to select  $A'$  if  $\alpha_H \cdot \bar{u}(A') + (1 - \alpha_H) \cdot \underline{u}(A') > \alpha_H \cdot \bar{u}(A'') + (1 - \alpha_H) \cdot \underline{u}(A'')$  and  $A''$  otherwise.

**Which value  $\alpha_H$  should we choose? An argument in favor of  $\alpha_H = 0.5$ .**  
Which value  $\alpha_H$  should we choose?

To answer this question, let us take an event  $E$  about which we know nothing. For a lottery  $L^+$  in which we get  $A_1$  if  $E$  and  $A_0$  otherwise, the utility interval is  $[0, 1]$ , thus, from a decision making viewpoint, this lottery should be equivalent to an event with utility  $\alpha_H \cdot 1 + (1 - \alpha_H) \cdot 0 = \alpha_H$ .

Similarly, for a lottery  $L^-$  in which we get  $A_0$  if  $E$  and  $A_1$  otherwise, the utility interval is  $[0, 1]$ , thus, this lottery should also be equivalent to an event with utility  $\alpha_H \cdot 1 + (1 - \alpha_H) \cdot 0 = \alpha_H$ .

We can now combine these two lotteries into a single complex lottery, in which we select either  $L^+$  or  $L^-$  with equal probability 0.5. Since  $L^+$  is equivalent to a lottery  $L(\alpha_H)$  with utility  $\alpha_H$  and  $L^-$  is also equivalent to a lottery

$L(\alpha_H)$  with utility  $\alpha_H$ , the complex lottery is equivalent to a lottery in which we select either  $L(\alpha_H)$  or  $L(\alpha_H)$  with equal probability 0.5, i.e., to  $L(\alpha_H)$ . Thus, the complex lottery has an equivalent utility  $\alpha_H$ .

On the other hand, no matter what is the event  $E$ , in the above complex lottery, we get  $A_1$  with probability 0.5 and  $A_0$  with probability 0.5. Thus, this complex lottery coincides with the lottery  $L(0.5)$  and thus, has utility 0.5. Thus, we conclude that  $\alpha_H = 0.5$ .

*Comment.* The fact that people with too optimistic attitude often make sub-optimal decisions is experimentally confirmed, e.g., in [42].

**Which action should we choose?** Suppose that an action has  $n$  possible outcomes  $S_1, \dots, S_n$ , with utilities

$$[\underline{u}(S_i), \bar{u}(S_i)],$$

and probabilities  $[\underline{p}_i, \bar{p}_i]$ . How do we then estimate the equivalent utility of this action?

We know that each alternative is equivalent to a simple lottery with utility  $u_i = \alpha_H \cdot \bar{u}(S_i) + (1 - \alpha_H) \cdot \underline{u}(S_i)$ , and that for each  $i$ , the  $i$ -th event is – from the viewpoint of decision making – equivalent to  $p_i = \alpha_H \cdot \bar{p}_i + (1 - \alpha_H) \cdot \underline{p}_i$ . Thus, from the viewpoint of decision making, this action is equivalent to a situation in which we get utility  $u_i$  with probability  $p_i$ . We know that the utility of such a situation is equal to  $\sum_{i=1}^n p_i \cdot u_i$ . Thus, the equivalent utility of the original action is equivalent to

$$\sum_{i=1}^n p_i \cdot u_i = \sum_{i=1}^n (\alpha_H \cdot \bar{p}_i + (1 - \alpha_H) \cdot \underline{p}_i) \cdot (\alpha_H \cdot \bar{u}(S_i) + (1 - \alpha_H) \cdot \underline{u}(S_i)).$$

*Comment.* One can easily see that if we replace the selected values  $A_0$  and  $A_1$  with  $A'_0$  and  $A'_1$ , so that the utilities change linearly  $u \rightarrow u' = a \cdot u + b$ , then the above equivalent utility  $u_{\text{equiv}}$  also changes according to the same linear transformation  $u'_{\text{equiv}} = a \cdot u_{\text{equiv}} + b$ .

**Discussion.** We started with the situation in which a decision maker cannot decide between  $A'$  and  $A''$ . In this case, it is possible that  $A'$  is better, and it is also possible that  $A''$  is better. In terms of interval-valued utilities  $[\underline{u}(A'), \bar{u}(A')]$  and  $[\underline{u}(A''), \bar{u}(A'')]$ , this means that:

- there exists values  $u(A') \in [\underline{u}(A'), \bar{u}(A')]$  and  $u(A'') \in [\underline{u}(A''), \bar{u}(A'')]$  for which  $u(A') > u(A'')$ , and
- there exists values  $u(A') \in [\underline{u}(A'), \bar{u}(A')]$  and  $u(A'') \in [\underline{u}(A''), \bar{u}(A'')]$  for which  $u(A') < u(A'')$ .

In this case, the above approach recommends selecting one of the alternatives  $A'$  and  $A''$ :

- we recommend to select  $A'$  if

$$\alpha_H \cdot \bar{u}(A') + (1 - \alpha_H) \cdot \underline{u}(A') \geq \alpha_H \cdot \bar{u}(A'') + (1 - \alpha_H) \cdot \underline{u}(A'');$$

- we recommend to select  $A''$  if

$$\alpha_H \cdot \bar{u}(A') + (1 - \alpha_H) \cdot \underline{u}(A') < \alpha_H \cdot \bar{u}(A'') + (1 - \alpha_H) \cdot \underline{u}(A'').$$

In this case, from the viewpoint of *descriptive* preference, we have uncertainty – we cannot decide between  $A'$  and  $A''$ . In this case, we make a recommendation. The recommended *prescriptive (normative)* preference will enable the user to make a good decision in a situation when this user is unsure which decision is better – this is exactly the type of situation in which user seek advise of specialists in decision making.

**Observation: the resulting decision depends on the level of detail.**

We make a decision in a situation when we do not know the exact values of the utilities and when we do not know the exact values of the corresponding probabilities. Clearly, if gain new information, the equivalent utility may change. For example, if we know nothing about an alternative  $A$ , then its utility is  $[0, 1]$  and thus, its equivalent utility is  $\alpha_H$ . Once we narrow down the utility of  $A$ , e.g., to the interval  $[0.5, 0.9]$ , we get a different equivalent utility  $\alpha_H \cdot 0.9 + (1 - \alpha_H) \cdot 0.5 = 0.5 + 0.4 \cdot \alpha_H$ . On this example, the fact that we have different utilities makes perfect sense.

However, there are other examples where the corresponding difference is not as intuitively clear. Let us consider a situation in which, with some probability  $p$ , we gain a utility  $u$ , and with the remaining probability  $1 - p$ , we gain utility 0. If we know the exact values of  $u$  and  $p$ , we can then compute the equivalent utility of this situation as the expected utility value  $p \cdot u + (1 - p) \cdot 0 = p \cdot u$ .

Suppose now that we only know the interval  $[\underline{u}, \bar{u}]$  of possible values of utility and the interval  $[\underline{p}, \bar{p}]$  of possible values of probability. Since the expression  $p \cdot u$  for the expected utility of this situation is an increasing function of both variables:

- the largest possible utility of this situation is attained when both  $p$  and  $u$  are the largest possible:  $u = \bar{u}$  and  $p = \bar{p}$ , and
- the smallest possible utility is attained when both  $p$  and  $u$  are the smallest possible:  $u = \underline{u}$  and  $p = \underline{p}$ .

In other words, the resulting amount of utility ranges from  $\underline{p} \cdot \underline{u}$  to  $\bar{p} \cdot \bar{u}$ .

If we know the structure of the situation, then, according to our derivation, this situation has an equivalent utility

$$u_k = (\alpha_H \cdot \bar{p} + (1 - \alpha_H) \cdot \underline{p}) \cdot (\alpha_H \cdot \bar{u} + (1 - \alpha_H) \cdot \underline{u})$$

(*k* for *know*). On the other hand, if we do not know the structure, if we only know that the resulting utility is from the interval  $[\underline{p} \cdot \underline{u}, \bar{p} \cdot \bar{u}]$ , then, according to the Hurwicz criterion, the equivalent utility is equal to

$$u_d = \alpha_H \cdot \bar{p} \cdot \bar{u} + (1 - \alpha_H) \cdot \underline{p} \cdot \underline{u}$$

(*d* for *don't know*). One can check that

$$\begin{aligned} u_d - u_k &= \\ \alpha_H \cdot \bar{p} \cdot \bar{u} + (1 - \alpha_H) \cdot \underline{p} \cdot \underline{u} - \alpha_H^2 \cdot \bar{p} \cdot \bar{u} - \alpha_H \cdot (1 - \alpha_H) \cdot (\underline{p} \cdot \bar{u} + \bar{p} \cdot \underline{u}) - (1 - \alpha_H)^2 \cdot \underline{p} \cdot \underline{u} &= \\ \alpha_H \cdot (1 - \alpha_H) \cdot \bar{p} \cdot \bar{u} + \alpha_H \cdot (1 - \alpha_H) \cdot \underline{p} \cdot \underline{u} - \alpha_H \cdot (1 - \alpha_H) \cdot (\underline{p} \cdot \bar{u} + \bar{p} \cdot \underline{u}) &= \\ \alpha_H \cdot (1 - \alpha_H) \cdot (\bar{p} - \underline{p}) \cdot (\bar{u} - \underline{u}). \end{aligned}$$

This difference is always positive, meaning that additional knowledge decreases the utility of the situation. (This is maybe what the Book of Ecclesiastes means by “For with much wisdom comes much sorrow”?)

*Comment.* A similar example has been recently described in [12].

## 5 From intervals to arbitrary sets

In the ideal case, we know the exact situation  $s$  in all the detail, and we can thus determine its utility  $u(s)$ . Realistically, we have an imprecise knowledge, so instead of a single situation  $s$ , we only know a *set*  $S$  of possible situations  $s$ . Thus, instead of a single value of the utility, we only know that the actual utility belongs to the set  $U = \{u(s) : s \in S\}$ . If this set  $S$  is an interval  $[\underline{u}, \bar{u}]$ , then we can use the above arguments to come up with its equivalent utility value  $\alpha_H \cdot \bar{u} + (1 - \alpha_H) \cdot \underline{u}$ .

What is  $U$  if it is not an interval? For example, we can have a 2-point set  $U = \{\underline{u}, \bar{u}\}$ . What is then the equivalent utility?

Let us first consider the case when the set  $U$  contains both its infimum  $\underline{u}$  and its supremum  $\bar{u}$ . The fact that we only know the set of possible values and have no other information means that *any* probability distribution on this set is possible (to be more precise, it is possible to have any probability distribution on the set of possible situations  $S$ , and this leads to the probability distribution on utilities). In particular, for each probability  $p$ , it is possible to have a distribution in which we have  $\bar{u}$  with probability  $p$  and  $\underline{u}$  with probability  $1 - p$ . For this distribution, the expected utility is equal to  $p \cdot \bar{u} + (1 - p) \cdot \underline{u}$ . When  $p$  goes from 0 to 1, these values fill the whole interval  $[\underline{u}, \bar{u}]$ . Thus, every value from this interval is the possible value of the expected utility. On the other hand, when  $u \in [\underline{u}, \bar{u}]$ , the expected value of the utility also belongs to this interval – no matter what the probability distribution. Thus, the set of all possible utility values is the whole interval  $[\underline{u}, \bar{u}]$  and so, the equivalent utility is equal to  $\alpha_H \cdot \bar{u} + (1 - \alpha_H) \cdot \underline{u}$ .

When the infimum and/or supremum are not in the set  $S$ , then the set  $S$  contains points as close to them as possible. Thus, the resulting set of possible

values of utility is as close as possible to the interval  $[\underline{u}, \bar{u}]$  – and so, it is reasonable to assume that the equivalent utility is as close to  $u_0 = \alpha_H \cdot \bar{u} + (1 - \alpha_H) \cdot \underline{u}$  as possible – i.e., coincides with this value  $u_0$ .

## 6 Beyond interval and set uncertainty: partial information about probabilities

**Formulation of the problem.** In addition to the interval  $\mathbf{x}$ , we may also have *partial* information about the probabilities of different values  $x \in \mathbf{x}$ . How can we describe this partial information?

An *exact* probability distribution can be described, e.g., by its cumulative distribution function (cdf)  $F(z) = \text{Prob}(x \leq z)$ . A *partial* information means that for each  $z$ , instead of knowing the exact value  $F(z)$ , we only know the bounds on  $F(z)$ , i.e., we only know the interval  $\mathbf{F}(z) = [F(z), \bar{F}(z)]$ . Such an interval-valued cdf is known as a *p-box*; see, e.g., [7, 32]. Once we know the p-box, we consider all possible distributions for which, for all  $z$ , we have  $F(z) \in \mathbf{F}(z)$ .

The problem is that there are many ways to represent a probability distribution, and each leads to a different way to represent partial information. Which of these ways should we choose?

**Which is the best way to describe the corresponding probabilistic uncertainty?** One of the main objectives of data processing is to make decisions. A standard way of making a decision is to select the action  $a$  for which the expected utility (gain) is the largest possible. This is where probabilities are used: in computing, for every possible action  $a$ , the corresponding expected utility. To be more precise, we usually know, for each action  $a$  and for each actual value of the (unknown) quantity  $x$ , the corresponding value of the utility  $u_a(x)$ . We must use the probability distribution for  $x$  to compute the expected value  $E[u_a(x)]$  of this utility.

In view of this application, the most useful characteristics of a probability distribution would be the ones which would enable us to compute the expected value  $E[u_a(x)]$  of different functions  $u_a(x)$ .

**Which representations are the most useful for this intended usage? General idea.** Which characteristics of a probability distribution are the most useful for computing mathematical expectations of different functions  $u_a(x)$ ? The answer to this question depends on the type of the function, i.e., on how the utility value  $u$  depends on the value  $x$  of the analyzed parameter.

**Smooth utility functions naturally lead to moments.** One natural case is when the utility function  $u_a(x)$  is smooth. We have already mentioned, in the previous text, that we usually know a (reasonably narrow) interval of possible values of  $x$ . So, to compute the expected value of  $u_a(x)$ , all we need to know is

how the function  $u_a(x)$  behaves on this narrow interval. Because the function is smooth, we can expand it into Taylor series. Because the interval is narrow, we can consider only linear and quadratic terms in this expansion and safely ignore higher-order terms:  $u_a(x) \approx c_0 + c_1 \cdot (x - x_0) + c_2 \cdot (x - x_0)^2$ , where  $x_0$  is a point inside the interval. Thus, we can approximate the expected value of this function by the expected value of the corresponding quadratic expression:  $E[u_a(x)] \approx E[c_0 + c_1 \cdot (x - x_0) + c_2 \cdot (x - x_0)^2]$ , i.e., by the following expression:  $E[u_a(x)] \approx c_0 + c_1 \cdot E[x - x_0] + c_2 \cdot E[(x - x_0)^2]$ . So, to compute the expectations of such utility functions, it is sufficient to know the first and second moments of the probability distribution.

In particular, if we use, as the point  $x_0$ , the average  $E[x]$ , the second moment turns into the variance of the original probability distribution. So, instead of the first and the second moments, we can use the mean  $E$  and the variance  $V$ .

**In decision making, non-smooth utility functions are common.** In decision making, not all dependencies are smooth. There is often a threshold  $x_0$  after which, say, a concentration of a certain chemical becomes dangerous.

This threshold sometimes comes from the detailed chemical and/or physical analysis. In this case, when we increase the value of this parameter, we see the drastic increase in effect and hence, the drastic change in utility value. Sometimes, this threshold simply comes from regulations. In this case, when we increase the value of this parameter past the threshold, there is no drastic increase in effects, but there is a drastic decrease of utility due to the necessity to pay fines, change technology, etc. In both cases, we have a utility function which experiences an abrupt decrease at a certain threshold value  $x_0$ .

**Non-smooth utility functions naturally lead to cumulative distribution functions (cdfs).** We want to be able to compute the expected value  $E[u_a(x)]$  of a function  $u_a(x)$  which

- changes smoothly until a certain value  $x_0$ ,
- then drops its value and continues smoothly for  $x > x_0$ .

We usually know the (reasonably narrow) interval which contains all possible values of  $x$ . Because the interval is narrow and the dependence before and after the threshold is smooth, the resulting change in  $u_a(x)$  before  $x_0$  and after  $x_0$  is much smaller than the change at  $x_0$ . Thus, with a reasonable accuracy, we can ignore the small changes before and after  $x_0$ , and assume that the function  $u_a(x)$  is equal to a constant  $u^+$  for  $x < x_0$ , and to some other constant  $u^- < u^+$  for  $x > x_0$ .

The simplest case is when  $u^+ = 1$  and  $u^- = 0$ . In this case, the desired expected value  $E[u_a^{(0)}(x)]$  coincides with the probability that  $x < x_0$ , i.e., with the corresponding value  $F(x_0)$  of the cumulative distribution function (cdf). A generic function  $u_a(x)$  of this type, with arbitrary values  $u^-$  and  $u^+$ , can be easily reduced to this simplest case, because, as one can easily check,  $u_a(x) = u^- + (u^+ - u^-) \cdot u^{(0)}(x)$  and hence,  $E[u_a(x)] = u^- + (u^+ - u^-) \cdot F(x_0)$ .

Thus, to be able to easily compute the expected values of all possible non-smooth utility functions, it is sufficient to know the values of the cdf  $F(x_0)$  for all possible  $x_0$ .

Describing the cdf is equivalent to describing the inverse *quantile* function – a function that assigns, to every possible probability  $p \in [0, 1]$ , the value  $x = x(p)$  for which  $F(x) = p$ . For example, the quantile corresponding to  $p = 0.5$  is the *median* of the probability distribution.

**Summarizing: which statistical characteristics we select.** Our analysis shows that the most appropriate characteristics are the moments and the values of the cdf (or, equivalently, the values of the quantiles).

*Comment.* How to estimate the values of the selected statistical characteristics? How to propagate these values via data processing? For answers to these questions, see [7, 32] and references therein.

## 7 What if we cannot even elicit interval-valued uncertainty: symmetry approach

**Case study.** In some situations, it is difficult to elicit even interval-valued utilities. As a case study, we consider the problem of selecting the best location for a meteorological tower.

In many applications involving meteorology and environmental sciences, it is important to measure fluxes of heat, water, carbon dioxide, methane and other trace gases that are exchanged within the atmospheric boundary layer. Air flow in this boundary layer consists of numerous rotating eddies, i.e., turbulent vortices of various sizes, with each eddy having horizontal and vertical components. To estimate the flow amount at a given location, we thus need to accurately measure wind speed (and direction), temperature, atmospheric pressure, gas concentration, etc., at different heights, and then process the resulting data. To perform these measurements, researchers build up vertical towers equipped with sensors at different heights; these tower are called *Eddy flux towers*.

When selecting a location for the Eddy flux tower, we have several criteria to satisfy; see, e.g., [2, 16].

- For example, the station should not be located too close to a road, so that the gas flux generated by the cars does not influence our measurements of atmospheric fluxes; in other words, the distance  $x_1$  to the road should be larger than a certain threshold  $t_1$ :  $x_1 > t_1$ , or  $y_1 \stackrel{\text{def}}{=} x_1 - t_1 > 0$ .
- Also, the inclination  $x_2$  at the station location should be smaller than a corresponding threshold  $t_2$ , because otherwise, the flux will be mostly determined by this inclination and will not be reflective of the atmospheric processes:  $x_2 < t_2$ , or  $y_2 \stackrel{\text{def}}{=} t_2 - x_2 > 0$ .



**General case.** In general, we have several such differences  $y_1, \dots, y_n$  all of which have to be non-negative. For each of the differences  $y_i$ , the larger its value, the better. Based on the above, our problem is a typical setting for *multi-criteria optimization*; see, e.g., [6, 37, 40].

*Practical problem: reminder.* We want to select the best location based on the values of the differences  $y_1, \dots, y_n$ . For each of the differences  $y_i$ , the larger its value, the better.

**Weighted average: a natural approach for solving multi-criterion optimization problems, and limitations of this approach.** The most widely used approach to multi-criteria optimization is *weighted average*, where we assign weights  $w_1, \dots, w_n > 0$  to different criteria  $y_i$  and select an alternative for which the weighted average  $w_1 \cdot y_1 + \dots + w_n \cdot y_n$  attains the largest possible value.

This approach has been used in many practical problems ranging from selecting the lunar landing sites for the Apollo missions (see, e.g., [3]) to selecting landfill sites (see, e.g., [10]).

In our problem, we have an additional requirement – that all the values  $y_i$  must be positive. Thus, we must only compare solutions with  $y_i > 0$  when selecting an alternative with the largest possible value of the weighted average.

In general, the weighted average approach often leads to reasonable solutions of the multi-criteria optimization problem. However, as we will show, in the presence of the additional positivity requirement, the weighted average approach is not fully satisfactory.

**A practical multi-criteria optimization must take into account that measurements are not absolutely accurate.** In many practical application of the multi-criterion optimization problem (in particular, in applications to optimal sensor placement), the values  $y_i$  come from measurements, and measurements are never absolutely accurate. The results  $\tilde{y}_i$  of the measurements are close to the actual (unknown) values  $y_i$  of the measured quantities, but they are not exactly equal to these values. If:

- we measure the values  $y_i$  with higher and higher accuracy and,
- based on the measurement results  $\tilde{y}_i$ , we conclude that the alternative  $y = (y_1, \dots, y_n)$  is better than some other alternative  $y' = (y'_1, \dots, y'_n)$ ,

then we expect that the actual alternative  $y$  is indeed either better than  $y'$  or at least of the same quality as  $y'$ . Otherwise, if we do not make this assumption, we will not be able to make any meaningful conclusions based on real-life (approximate) measurements.

**The above natural requirement is not always satisfied for weighted average.** Let us show that for the weighted average, this “continuity” requirement is not satisfied even in the simplest case when we have only two criteria  $y_1$  and  $y_2$ . Indeed, let  $w_1 > 0$  and  $w_2 > 0$  be the weights corresponding to these two criteria. Then, the resulting strict preference relation  $\succ$  has the following properties:

- if  $y_1 > 0$ ,  $y_2 > 0$ ,  $y'_1 > 0$ , and  $y'_2 > 0$ , and  $w_1 \cdot y'_1 + w_2 \cdot y'_2 > w_1 \cdot y_1 + w_2 \cdot y_2$ , then

$$y' = (y'_1, y'_2) \succ y = (y_1, y_2); \quad (1)$$

- if  $y_1 > 0$ ,  $y_2 > 0$ , and at least one of the values  $y'_1$  and  $y'_2$  is non-positive, then

$$y = (y_1, y_2) \succ y' = (y'_1, y'_2). \quad (2)$$

Let us consider, for every  $\varepsilon > 0$ , the tuple  $y'(\varepsilon) \stackrel{\text{def}}{=} \left( \varepsilon, 1 + \frac{w_1}{w_2} \right)$ , with  $y'_1(\varepsilon) = \varepsilon$  and  $y'_2(\varepsilon) = 1 + \frac{w_1}{w_2}$ , and also the comparison tuple  $y = (1, 1)$ . In this case, for every  $\varepsilon > 0$ , we have

$$w_1 \cdot y'_1(\varepsilon) + w_2 \cdot y'_2(\varepsilon) = w_1 \cdot \varepsilon + w_2 + w_2 \cdot \frac{w_1}{w_2} = w_1 \cdot (1 + \varepsilon) + w_2 \quad (3)$$

and

$$w_1 \cdot y_1 + w_2 \cdot y_2 = w_1 + w_2, \quad (4)$$

hence  $y'(\varepsilon) \succ y$ . However, in the limit  $\varepsilon \rightarrow 0$ , we have  $y'(0) = \left( 0, 1 + \frac{w_1}{w_2} \right)$ , with  $y'_1(0) = 0$  and thus,  $y'(0) \prec y$ .

### **Towards a more adequate approach to multi-criterion optimization.**

We want to be able to compare different alternatives.

Each alternative is characterized by a tuple of  $n$  values  $y = (y_1, \dots, y_n)$ , and only alternatives for which all the values  $y_i$  are positive are allowed. Thus, from the mathematical viewpoint, the set of all alternatives is the set  $(R^+)^n$  of all the tuples of positive numbers.

For each two alternatives  $y$  and  $y'$ , we want to tell whether  $y$  is better than  $y'$  (we will denote it by  $y \succ y'$  or  $y' \prec y$ ), or  $y'$  is better than  $y$  ( $y' \succ y$ ), or  $y$  and  $y'$  are equally good ( $y' \sim y$ ). These relations must satisfy natural properties. For example, if  $y$  is better than  $y'$  and  $y'$  is better than  $y''$ , then  $y$  is better than  $y''$ . In other words, the relation  $\succ$  must be transitive. Similarly, the relation  $\sim$  must be transitive, symmetric, and reflexive ( $y \sim y$ ), i.e., in mathematical terms, an *equivalence relation*.

So, we want to define a pair of relations  $\succ$  and  $\sim$  such that  $\succ$  is transitive,  $\sim$  is an equivalence relation, and for every  $y$  and  $y'$ , one and only one of the following relations hold:  $y \succ y'$ ,  $y' \succ y$ , or  $y \sim y'$ .

It is also reasonable to require that if each criterion is better, then the alternative is better as well, i.e., that if  $y_i > y'_i$  for all  $i$ , then  $y \succ y'$ .

*Comment.* Pairs of relations of the above type can be alternatively characterized by a *pre-ordering* relation

$$y' \succeq y \Leftrightarrow (y' \succ y \vee y' \sim y). \quad (5)$$

This pre-ordering relation must be transitive and – in our case – total (i.e., for every  $y$  and  $y'$ , we have  $y \succeq y' \vee y' \succeq y$ ). Once we know the pre-ordering relation  $\succeq$ , we can reconstruct  $\succ$  and  $\sim$  as follows:

$$y' \succ y \Leftrightarrow (y' \succeq y \ \& \ y \not\succeq y'); \quad (6)$$

$$y' \sim y \Leftrightarrow (y' \succeq y \ \& \ y \succeq y'). \quad (7)$$

**Scale invariance: motivation.** In general, the quantities  $y_i$  describe completely different physical notions, measured in completely different units. In our meteorological case, some of these values are wind velocities measured in meters per second, or in kilometers per hour, or in miles per hour. Other values are elevations described in meters, in kilometers, or in feet, etc. Each of these quantities can be described in many different units. A priori, we do not know which units match each other, so it is reasonable to assume that the units used for measuring different quantities may not be exactly matched.

It is therefore reasonable to require that the relations  $\succ$  and  $\sim$  between the two alternatives  $y = (y_1, \dots, y_n)$  and  $y' = (y'_1, \dots, y'_n)$  do not change if we simply change the units in which we measure each of the corresponding  $n$  quantities.

*Comment.* The importance of such invariance is well known in measurements theory, starting with the pioneering work of S. S. Stevens ([39]); see also the classical books [33] and [25] (especially Chapter 22), where this invariance is also called *meaningfulness*.

**Scale invariance: towards a precise description.** When we replace a unit in which we measure a certain quantity  $q$  by a new measuring unit which is  $\lambda > 0$  times smaller, then the numerical values of this quantity increase by a factor of  $\lambda$ , i.e.,  $q \rightarrow \lambda \cdot q$ . For example, 1 cm is  $\lambda = 100$  times smaller than 1 m, so the length  $q = 2$  m, when measured in cm, becomes  $\lambda \cdot q = 2 \cdot 100 = 200$  cm.

Let  $\lambda_i$  denote the ratio of the old to the new units corresponding to the  $i$ -th quantity. Then, the quantity that had the value  $y_i$  in the old units will be described by a numerical value  $\lambda_i \cdot y_i$  in the new units. Therefore, scale-invariance means that for all  $y, y' \in (R^+)^n$  and for all  $\lambda_i > 0$ , we have

$$y' = (y'_1, \dots, y'_n) \succ y = (y_1, \dots, y_n) \Rightarrow (\lambda_1 \cdot y'_1, \dots, \lambda_n \cdot y'_n) \succ (\lambda_1 \cdot y_1, \dots, \lambda_n \cdot y_n)$$

and

$$y' = (y'_1, \dots, y'_n) \sim y = (y_1, \dots, y_n) \Rightarrow (\lambda_1 \cdot y'_1, \dots, \lambda_n \cdot y'_n) \sim (\lambda_1 \cdot y_1, \dots, \lambda_n \cdot y_n).$$

*Comment.* In general, in measurements, in addition to changing the unit, we can also change the starting point. However, for the differences  $y_i$ , the starting point is fixed by the fact that 0 corresponds to the threshold value. So, in our case, only changing a measuring unit (= scaling) makes sense.

**Continuity.** As we have mentioned in the previous section, we also want to require that the relations  $\succ$  and  $\sim$  are *continuous* in the following sense: if  $y'(\varepsilon) \succeq y(\varepsilon)$  for every  $\varepsilon$ , then in the limit, when  $y'(\varepsilon) \rightarrow y'(0)$  and  $y(\varepsilon) \rightarrow y(0)$  (in the sense of normal convergence in  $R^n$ ), we should have  $y'(0) \succeq y(0)$ .

**The main result.** Let us now describe our requirements in precise terms.

**Definition 2.** By a total pre-ordering relation on a set  $Y$ , we mean a pair of a transitive relation  $\succ$  and an equivalence relation  $\sim$  for which, for every  $y, y' \in Y$ , one and only one of the following relations hold:  $y \succ y'$ ,  $y' \succ y$ , or  $y \sim y'$ .

*Comment.* We will denote  $y \succeq y' \stackrel{\text{def}}{=} (y \succ y' \vee y \sim y')$ .

**Definition 3.** We say that a total pre-ordering is non-trivial if there exist  $y$  and  $y'$  for which  $y' \succ y$ .

*Comment.* This definition excludes the trivial pre-ordering in which every two tuples are equivalent to each other.

**Definition 4.** We say that a total pre-ordering relation on the set  $(R^+)^n$  is:

- monotonic if  $y'_i > y_i$  for all  $i$  implies  $y' \succ y$ ;
- scale-invariant if for all  $\lambda_i > 0$ :

- $(y'_1, \dots, y'_n) \succ y = (y_1, \dots, y_n)$  implies

$$(\lambda_1 \cdot y'_1, \dots, \lambda_n \cdot y'_n) \succ (\lambda_1 \cdot y_1, \dots, \lambda_n \cdot y_n), \quad (8)$$

and

- $(y'_1, \dots, y'_n) \sim y = (y_1, \dots, y_n)$  implies

$$(\lambda_1 \cdot y'_1, \dots, \lambda_n \cdot y'_n) \sim (\lambda_1 \cdot y_1, \dots, \lambda_n \cdot y_n). \quad (9)$$

- continuous if whenever we have a sequence  $y^{(k)}$  of tuples for which  $y^{(k)} \succeq y'$  for some tuple  $y'$ , and the sequence  $y^{(k)}$  tends to a limit  $y$ , then  $y \succeq y'$ .

**Theorem.** [19] Every non-trivial monotonic scale-invariant continuous total pre-ordering relation on  $(R^+)^n$  has the following form:

$$y' = (y'_1, \dots, y'_n) \succ y = (y_1, \dots, y_n) \Leftrightarrow \prod_{i=1}^n (y'_i)^{\alpha_i} > \prod_{i=1}^n y_i^{\alpha_i}; \quad (10)$$

$$y' = (y'_1, \dots, y'_n) \sim y = (y_1, \dots, y_n) \Leftrightarrow \prod_{i=1}^n (y'_i)^{\alpha_i} = \prod_{i=1}^n y_i^{\alpha_i}, \quad (11)$$

for some constants  $\alpha_i > 0$ .

*Comment.* In other words, for every non-trivial monotonic scale-invariant continuous total pre-ordering relation on  $(R^+)^n$ , there exist values  $\alpha_1 > 0, \dots, \alpha_n > 0$  for which the above equivalence hold. Vice versa, for each set of values  $\alpha_1 > 0, \dots, \alpha_n > 0$ , the above formulas define a monotonic scale-invariant continuous pre-ordering relation on  $(R^+)^n$ .

For reader's convenience, the proof of the main result is presented in an Appendix.

It is worth mentioning that the resulting relation coincides with Cobb-Douglas production (utility) function [4, 43] and with the asymmetric version (see, e.g., [35]) of the bargaining solution proposed by the Nobelist John Nash (see next section).

**Applications.** We have applied this approach to selecting a site for the Eddy tower that we built at Jornada Experimental Range, a study site in the northern Chihuahuan Desert; see, e.g., [17, 18]. In this applications, the parameters  $y_i$  have already been identified in the previous research; see, e.g., [2].

The values  $\alpha_i$  were selected based on the information provided by experts, who supplied us with pairs of (approximately) equally good (or equally bad) designs  $y$  and  $y'$  with different combinations of the parameters  $y_i$ . Each resulting condition  $\prod_{i=1}^n y_i^{\alpha_i} = \prod_{i=1}^n (y'_i)^{\alpha_i}$  can be equivalently described, after taking

logarithms of both sides, as a linear equation  $\sum_{i=1}^n \alpha_i \cdot \ln(y_i) = \sum_{i=1}^n \alpha_i \cdot \ln(y'_i)$ .

By solving this system of linear equations, we found the values  $\alpha_i$  that reflect the expert opinion on the efficiency of Eddy towers.

A similar symmetry-based approach was used to design a network of radiotelesopes [23].

*Comment.* The above equations determine  $\alpha_i$  modulo a multiplicative constant: if we multiply all the values  $\alpha_i$  by the same constant, the equations remain valid. To avoid this non-uniqueness, we used normalized values of  $\alpha_i$ , i.e., values that satisfy the additional normalizing equation  $\sum_{i=1}^n \alpha_i = 1$ .

## 8 Group decision making

**Need for group decision making.** In many practical situations, several people are affected by the planned decision. In such situations, we need to take into account preferences of all the participating agents.

For each participant  $P_i$ , we can determine the utility  $u_{ij} \stackrel{\text{def}}{=} u_i(A_j)$  of all the alternatives  $A_1, \dots, A_m$ . How to transform these utilities into a reasonable group decision rule?

**Nash's bargaining solution.** The answer to this question was, in effect, provided by a future Nobelist John Nash who, in [29, 30], has shown that under reasonable assumptions like symmetry, independence from irrelevant alternatives, and *scale invariance* (i.e., invariance under replacing the original utility function  $u_i(A)$  with an equivalent function  $a \cdot u_i(A)$ ), the only group decision rule is selecting an alternative  $A$  for which the product

$$u(A) \stackrel{\text{def}}{=} \prod_{i=1}^n u_i(A)$$

is the largest possible; see also [26, 28].

Here, the utility functions must be scaled in such a way that the “status quo” situation  $A^{(0)}$  is assigned the utility 0. This re-scaling can be achieved, e.g., by replacing the original utility values  $u_i(A)$  with re-scaled values  $u'_i(A) \stackrel{\text{def}}{=} u_i(A) - u_i(A^{(0)})$ .

**Multi-agent decision making under interval uncertainty.** What if we do not know the exact values of utility, we only know intervals  $[\underline{u}_i(A), \bar{u}_i(A)]$ ? In this case, the first idea is to find all  $A_0$  which can be Nash-optimal, i.e., for which  $\bar{u}(A_0) \geq \max_A \underline{u}(A)$ , where

$$\underline{u}(A) \stackrel{\text{def}}{=} \prod_{i=1}^n \underline{u}_i(A) \text{ and } \bar{u}(A) \stackrel{\text{def}}{=} \prod_{i=1}^n \bar{u}_i(A).$$

If we want to select a single alternative, then we should maximize  $u^{\text{equiv}}(A) \stackrel{\text{def}}{=} \prod_{i=1}^n u_i^{\text{equiv}}(A)$ , where  $u_i^{\text{equiv}}(A)$  are values obtained by using Hurwicz optimism-pessimism criterion.

*Comment.* An interesting aspect of this problem is that sometimes, we have a conflict situation; this happens, for example, in security situations. In such situations, only partial results are known; see, e.g., [20].

## 9 Beyond optimization

**Need to go beyond optimization.** While optimization problems are ubiquitous, sometimes, we need to go beyond optimization: e.g., we need to make sure that the system is *controllable* for all disturbances within a given range.

In control situations, the desired value  $z$  depends both on the variables the variables that we can select (control variables)  $u = (u_1, \dots, u_m)$  and on the variables  $x = (x_1, \dots, x_n)$  describing the changing state of the world:  $z = f(x, u)$ . For each control variable  $u_j$ , we know the range  $U_j$  within which we can select its value, and for each variable  $x_i$ , we know the range  $X_i$  of its possible

values. We want to find a range  $Z$  for which, for every state of the world  $x_i \in X_i$ , we can get  $z \in Z$  by selecting appropriate control values  $u_j \in U_j$ :

$$\forall x \exists u (z = f(x, u) \in Z).$$

**Interval computations: reminder.** Interval computations [27] can be viewed as a degenerate case of this control problem in which there are no controls at all. In this case:

- we know the intervals  $X_1, \dots, X_n$  containing  $x_1, \dots, x_n$ ;
- we know that a quantity  $z$  depends on  $x$ :  $z = f(x)$ ;
- we want to find the range  $Z$  of possible values of  $z$ :

$$Z = \left[ \min_{x \in X} f(x), \max_{x \in X} f(x) \right].$$

In logical terms, we want to make sure that  $\forall x (z = f(x) \in Z)$ .

**Reformulation in logical terms – of modal intervals.** In the general control case, we want to make sure that  $\forall x \in X \exists u \in U (f(x, u) \in Z)$ . There is a logical difference between intervals  $X$  and  $U$ : the property  $f(x, u) \in Z$  must hold

- for all possible values  $x_i \in X_i$ , but
- for some values  $u_j \in U_j$ .

We can thus consider pairs of intervals and quantifiers (*modal intervals* [11]):

- each original interval  $X_i$  is a pair  $\langle X_i, \forall \rangle$ , while
- controlled interval is a pair  $\langle U_j, \exists \rangle$ .

We can then treat the resulting interval  $Z$  as the “range” defined over such modal intervals:

$$Z = f(\langle X_1, \forall \rangle, \dots, \langle X_n, \forall \rangle, \langle U_1, \exists \rangle, \dots, \langle U_m, \exists \rangle).$$

**Even further beyond optimization.** In more complex situations, we need to go beyond control. For example, in the presence of an adversary, we want to make a decision  $x$  such that:

- for every possible reaction  $y$  of an adversary,
- we will be able to make a next decision  $x'$  (depending on  $y$ )
- so that after every possible next decision  $y'$  of an adversary,
- the resulting state  $s(x, y, x', y')$  will be in the desired set:

$$\forall y \exists x \forall y' (s(x, y, x', y') \in S).$$

In this case, we arrive at general quantifier classes described, e.g., in [38].

**Acknowledgments.** This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, by Grant 1 T36 GM078000-01 from the National Institutes of Health, and by a grant on F-transforms from the Office of Naval Research.

The authors is greatly thankful to the anonymous referees for valuable suggestions.

## References

- [1] R. Aliev, O. Huseynov, and V. Kreinovich, “Decision Making under Interval and Fuzzy Uncertainty: Towards an Operational Approach”, *Proceedings of the Tenth International Conference on Application of Fuzzy Systems and Soft Computing ICAFS’2012*, Lisbon, Portugal, August 29–30, 2012.
- [2] M. Aubinet, T. Vesala, and D. Papale (eds.), *Eddy Covariance – A Practical Guide to Measurement and Data Analysis*, Springer, Dordrecht, Hiedelberg, London, New York, 2012.
- [3] A. B. Binder and D. L. Roberts, *Criteria for Lunar Site Selection*, Report No. P-30, NASA Appollo Lunar Exploration Office and Illinois Institute of Technology Research Institute, Chicago, Illinois, January 1970.
- [4] C. W. Cobb and P. H. Douglas, “A Theory of Production”, *American Economic Review*, 1928, Vol. 18, Supplement, pp. 139–165.
- [5] B. de Finetti, *Theory of Probability*, Wiley, New York, Vol. 1, 1974; Vol. 2, 1975.
- [6] M. Ehrgott and X. Gandibleux (eds.), *Multiple Criteria Optimization: State of the Art Annotated Bibliographic Surveys*, Springer Verlag, Berlin-Heidelberg-New York, 2002.
- [7] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, and L. Ginzburg, *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty*, Sandia National Laboratories, 2007, Publ. 2007-0939.
- [8] P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.
- [9] P. C. Fishburn, *Nonlinear Preference and Utility Theory*, Johns Hopkins University Press, Baltimore, 1988.
- [10] I. Fountoulis, D. Mariokalos, E. Spyridonos, and E. Andreakis, “Geological criteria and methodology for landfill sites selection”, In: *Proceedings of the 8th International Conference on Environmental Science and Technology*, Lemnos Island, Greece, September 8–10, 2003, pp. 200–207.



- [11] E. Gardeñes et al., Modal intervals, *Reliable Computing*, 2001, Vol. 7, pp. 77–111.
- [12] P. H. Giang, “Decision with Dempster-Shafer belief functions: Decision under ignorance and sequential consistency”, *International Journal of Approximate Reasoning*, 2012, Vol. 53, pp. 38–53.
- [13] P. Guo and H. Tanaka, “Decision making with interval probabilities”, *European Journal of Operational Research*, 2010, Vol. 203, pp. 444–454.
- [14] P. Guo and Y. Wang, “Eliciting dual interval probabilities from interval comparison matrices”, *Information Sciences*, 2012, Vol. 190, pp. 17–26.
- [15] L. Hurwicz, *Optimality Criteria for Decision Making Under Ignorance*, Cowles Commission Discussion Paper, Statistics, No. 370, 1951.
- [16] A. Jaimes, *A cyber-tool to optimize site selection for establishing an eddy covariance and robotic tram system at the Jornada Experimental Range*, University of Texas at El Paso, December 2008.
- [17] A. Jaimes, J. Herrera, L. González, G. Ramírez, C. Laney, D. Browning, D. Peters, M. Litvak, and C. Tweedie, “Towards a multiscale approach to link climate, NEE and optical properties from a flux tower, robotic tram system that measures hyperspectral reflectance, phenocams, phenostations and a sensor network in desert a shrubland”, In: *Proceedings of the FLUXNET and Remote Sensing Open Workshop: Towards Upscaling Flux Information from Towers to the Globe FLUXNET-SpecNet’2011*, Berkeley, California, June 7–9, 2011.
- [18] A. Jaimes, L. Salayandia, and I. Gallegos, “New Cyber Infrastructure for Studying Land-Atmosphere Interactions Using Eddy Covariance Techniques”, In: *Abstracts of the 2010 Fall Meeting American Geophysical Union AGU’2010*, San Francisco, California, December 12–18, 2010.
- [19] A. Jaimes, C. Tweedie, V. Kreinovich, and M. Ceberio, Scale-invariant approach to multi-criterion optimization under uncertainty, with applications to optimal sensor placement, in particular, to sensor placement in environmental research, *International Journal of Reliability and Safety*, 2012, Vol. 6, No. 1–3, pp. 188–203.
- [20] C. Kiekintveld and V. Kreinovich, “Efficient approximation for security games with interval uncertainty”, *Proceedings of the AAAI Spring Symposium on Game Theory for Security, Sustainability, and Health GTSSH’2012*, Stanford, March 26–28, 2012.
- [21] E. Kintisch, “Loss of carbon observatory highlights gaps in data”, *Science*, 2009, Vol. 323, pp. 1276–1277.

- [22] V. Kreinovich, “Decision making under interval uncertainty”, *Abstracts of the 15th GAMM – IMACS International Symposium on Scientific Computing, Computer Arithmetic, and Verified Numerical Computation SCAN’2012*, Novosibirsk, Russia, September 23–29, 2012.
- [23] V. Kreinovich, S. A. Starks, D. Iourinski, O. Kosheleva, and A. Finkelstein, “Open-ended configurations of radio telescopes: a geometrical analysis”, *Geoinformatics*, 2003, Vol. 13, No. 2, pp. 79–85.
- [24] H. E. Kyburg and H. E. Smokler (eds), *Studies in Subjective Probability*, Wiley, New York, 1964.
- [25] R. D. Luce, D. H. Krantz, P. Suppes, and A. Tversky, *Foundations of Measurement, Vol. 3, Representation, Axiomatization, and Invariance*, Academic Press, San Diego, California, 1990.
- [26] R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.
- [27] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, Pennsylvania, 2009.
- [28] R. B. Myerson, *Game theory. Analysis of conflict*, Harvard University Press, Cambridge, MA, 1991.
- [29] J. F. Nash, “The bargaining problem”, *Econometrica*, 1950, Vol. 28, pp. 155–162.
- [30] J. Nash, “Two-Person Cooperative Games”, *Econometrica*, 1953, Vol. 21, pp. 128–140.
- [31] H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.
- [32] H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, 2012.
- [33] J. Pfanzagl, *Theory of Measurement*, John Wiley, New York, 1968.
- [34] H. Raiffa, *Decision Analysis*, McGraw-Hill, Columbus, Ohio, 1997.
- [35] A. Roth, *Axiomatic Models of Bargaining*, Springer-Verlag, Berlin, 1979.
- [36] L. J. Savage, *The Foundations of Statistics*, New York, Wiley, 1954.
- [37] Y. Sawaragi, H. Nakayama, and T. Tanino, *Theory of Multiobjective Optimization*, Academic Press, Orlando, Florida, 1985.
- [38] S. P. Shary, “A New Technique in Systems Analysis under Interval Uncertainty and Ambiguity”, *Reliable Computing*, 2002, Vol. 8, pp. 321–418.

- [39] S. S. Stevens, “On the theory of scales of measurement”, *Science*, 1946, Vol. 103, pp. 677–680.
- [40] R. E. Steuer, *Multiple Criteria Optimization: Theory, Computations, and Application*, John Wiley & Sons, New York, 1986.
- [41] M. C. M. Troffaes, “Decision making under uncertainty using imprecise probabilities”, *International Journal of Approximate Reasoning*, 2007, Vol. 45, pp. 7–29.
- [42] B. von Helversen and R. Mata, “Losing a dime with a satisfied mind: positive affect predicts less search in sequential decision making”, *Psychology and Aging*, 2012, to appear.
- [43] C. Walsh, *Monetary Theory and Policy*, MIT Press, Cambridge, Massachusetts, 2003.

## A Proof of the Theorem

1°. Due to scale-invariance (9), for every  $y_1, \dots, y_n, y'_1, \dots, y'_n$ , we can take  $\lambda_i = \frac{1}{y_i}$  and conclude that

$$(y'_1, \dots, y'_n) \sim (y_1, \dots, y_n) \Leftrightarrow \left( \frac{y'_1}{y_1}, \dots, \frac{y'_n}{y_n} \right) \sim (1, \dots, 1). \quad (12)$$

Thus, to describe the equivalence relation  $\sim$ , it is sufficient to describe the set of all the vectors  $z = (z_1, \dots, z_n)$  for which  $z \sim (1, \dots, 1)$ . Similarly,

$$(y'_1, \dots, y'_n) \succ (y_1, \dots, y_n) \Leftrightarrow \left( \frac{y'_1}{y_1}, \dots, \frac{y'_n}{y_n} \right) \succ (1, \dots, 1). \quad (13)$$

So, to describe the ordering relation  $\succ$ , it is sufficient to describe the set of all the vectors  $z = (z_1, \dots, z_n)$  for which  $z \succ (1, \dots, 1)$ .

Alternatively, we can take  $\lambda_i = \frac{1}{y'_i}$  and conclude that

$$(y'_1, \dots, y'_n) \succ (y_1, \dots, y_n) \Leftrightarrow (1, \dots, 1) \succ \left( \frac{y_1}{y'_1}, \dots, \frac{y_n}{y'_n} \right). \quad (14)$$

Thus, it is also sufficient to describe the set of all the vectors  $z = (z_1, \dots, z_n)$  for which  $(1, \dots, 1) \succ z$ .

2°. The above equivalence involves division. To simplify the description, we can take into account that in the logarithmic space, division becomes a simple difference:  $\ln\left(\frac{y'_i}{y_i}\right) = \ln(y'_i) - \ln(y_i)$ . To use this simplification, let us consider the logarithms  $Y_i \stackrel{\text{def}}{=} \ln(y_i)$  of different values. In terms of these logarithms, the original values can be reconstructed as  $y_i = \exp(Y_i)$ . In terms of these logarithms, we thus need to consider:

- the set  $S_{\sim}$  of all the tuples  $Z = (Z_1, \dots, Z_n)$  for which

$$z = (\exp(Z_1), \dots, \exp(Z_n)) \sim (1, \dots, 1), \quad (15)$$

and

- the set  $S_{\succ}$  of all the tuples  $Z = (Z_1, \dots, Z_n)$  for which

$$z = (\exp(Z_1), \dots, \exp(Z_n)) \succ (1, \dots, 1). \quad (16)$$

We will also consider the set  $S_{\prec}$  of all the tuples  $Z = (Z_1, \dots, Z_n)$  for which

$$(1, \dots, 1) \succ z = (\exp(Z_1), \dots, \exp(Z_n)). \quad (17)$$

Since the pre-ordering relation is total, for every tuple  $z$ ,

- either  $z \sim (1, \dots, 1)$ ,
- or  $z \succ (1, \dots, 1)$ ,
- or  $(1, \dots, 1) \succ z$ .

In particular, this is true for  $z = (\exp(Z_1), \dots, \exp(Z_n))$ . Thus, for every tuple  $Z$ , either  $Z \in S_{\sim}$  or  $Z \in S_{\succ}$  or  $Z \in S_{\prec}$ .

3°. Let us prove that the set  $S_{\sim}$  is closed under addition, i.e., that if the tuples  $Z = (Z_1, \dots, Z_n)$  and  $Z' = (Z'_1, \dots, Z'_n)$  belong to the set  $S_{\sim}$ , then their component-wise sum

$$Z + Z' = (Z_1 + Z'_1, \dots, Z_n + Z'_n) \quad (18)$$

also belongs to the set  $S_{\sim}$ .

Indeed, by definition (15) of the set  $S_{\sim}$ , the condition  $Z \in S_{\sim}$  means that

$$(\exp(Z_1), \dots, \exp(Z_n)) \sim (1, \dots, 1). \quad (19)$$

Using scale-invariance (9) with  $\lambda_i = \exp(Z'_i)$ , we conclude that

$$(\exp(Z_1) \cdot \exp(Z'_1), \dots, \exp(Z_n) \cdot \exp(Z'_n)) \sim (\exp(Z'_1), \dots, \exp(Z'_n)). \quad (20)$$

On the other hand, the condition  $Z' \in S_{\sim}$  means that

$$(\exp(Z'_1), \dots, \exp(Z'_n)) \sim (1, \dots, 1). \quad (21)$$

Thus, due to transitivity of the equivalence relation  $\sim$ , we conclude that

$$(\exp(Z_1) \cdot \exp(Z'_1), \dots, \exp(Z_n) \cdot \exp(Z'_n)) \sim (1, \dots, 1). \quad (22)$$

Since for every  $i$ , we have  $\exp(Z_i) \cdot \exp(Z'_i) = \exp(Z_i + Z'_i)$ , we thus conclude that

$$(\exp(Z_1 + Z'_1), \dots, \exp(Z_n + Z'_n)) \sim (1, \dots, 1). \quad (23)$$

By definition (15) of the set  $S_{\sim}$ , this means that the tuple  $Z + Z'$  belongs to the set  $S_{\sim}$ .

4°. Similarly, we can prove that the set  $S_{\succ}$  is closed under addition, i.e., that if the tuples  $Z = (Z_1, \dots, Z_n)$  and  $Z' = (Z'_1, \dots, Z'_n)$  belong to the set  $S_{\succ}$ , then their component-wise sum

$$Z + Z' = (Z_1 + Z'_1, \dots, Z_n + Z'_n) \quad (24)$$

also belongs to the set  $S_{\succ}$ .

Indeed, by definition (16) of the set  $S_{\succ}$ , the condition  $Z \in S_{\succ}$  means that

$$(\exp(Z_1), \dots, \exp(Z_n)) \succ (1, \dots, 1). \quad (25)$$

Using scale-invariance (8) with  $\lambda_i = \exp(Z'_i)$ , we conclude that

$$(\exp(Z_1) \cdot \exp(Z'_1), \dots, \exp(Z_n) \cdot \exp(Z'_n)) \succ (\exp(Z'_1), \dots, \exp(Z'_n)). \quad (26)$$

On the other hand, the condition  $Z' \in S_{\succ}$  means that

$$(\exp(Z'_1), \dots, \exp(Z'_n)) \succ (1, \dots, 1). \quad (27)$$

Thus, due to transitivity of the strict preference relation  $\succ$ , we conclude that

$$(\exp(Z_1) \cdot \exp(Z'_1), \dots, \exp(Z_n) \cdot \exp(Z'_n)) \succ (1, \dots, 1). \quad (28)$$

Since for every  $i$ , we have  $\exp(Z_i) \cdot \exp(Z'_i) = \exp(Z_i + Z'_i)$ , we thus conclude that

$$(\exp(Z_1 + Z'_1), \dots, \exp(Z_n + Z'_n)) \succ (1, \dots, 1). \quad (29)$$

By definition (16) of the set  $S_{\succ}$ , this means that the tuple  $Z + Z'$  belongs to the set  $S_{\succ}$ .

5°. A similar argument shows that the set  $S_{\prec}$  is closed under addition, i.e., that if the tuples  $Z = (Z_1, \dots, Z_n)$  and  $Z' = (Z'_1, \dots, Z'_n)$  belong to the set  $S_{\prec}$ , then their component-wise sum

$$Z + Z' = (Z_1 + Z'_1, \dots, Z_n + Z'_n) \quad (30)$$

also belongs to the set  $S_{\prec}$ .

6°. Let us now prove that the set  $S_{\sim}$  is closed under the “unary minus” operation, i.e., that if  $Z = (Z_1, \dots, Z_n) \in S_{\sim}$ , then  $-Z \stackrel{\text{def}}{=} (-Z_1, \dots, -Z_n)$  also belongs to  $S_{\sim}$ .

Indeed,  $Z \in S_{\sim}$  means that

$$(\exp(Z_1), \dots, \exp(Z_n)) \sim (1, \dots, 1). \quad (31)$$

Using scale-invariance (9) with  $\lambda_i = \exp(-Z_i) = \frac{1}{\exp(Z_i)}$ , we conclude that

$$(1, \dots, 1) \sim (\exp(-Z_1), \dots, \exp(-Z_n)), \quad (32)$$

i.e., that  $-Z \in S_{\sim}$ .

7°. Let us prove that if  $Z = (Z_1, \dots, Z_n) \in S_{\succ}$ , then  $-Z \stackrel{\text{def}}{=} (-Z_1, \dots, -Z_n)$  belongs to  $S_{\prec}$ .

Indeed,  $Z \in S_{\succ}$  means that

$$(\exp(Z_1), \dots, \exp(Z_n)) \succ (1, \dots, 1). \quad (33)$$

Using scale-invariance (8) with  $\lambda_i = \exp(-Z_i) = \frac{1}{\exp(Z_i)}$ , we conclude that

$$1, \dots, 1 \succ (\exp(-Z_1), \dots, \exp(-Z_n)), \quad (34)$$

i.e., that  $-Z \in S_{\prec}$ .

Similarly, we can show that if  $Z \in S_{\prec}$ , then  $-Z \in S_{\succ}$ .

8°. From Part 3 of this proof, it now follows that if  $Z = (Z_1, \dots, Z_n) \in S_{\sim}$ , then  $Z + Z \in S_{\sim}$ , then that  $Z + (Z + Z) \in S_{\sim}$ , etc., i.e., that for every positive integer  $p$ , the tuple

$$p \cdot Z = (p \cdot Z_1, \dots, p \cdot Z_n) \quad (35)$$

also belongs to the set  $S_{\sim}$ .

By using Part 6 of this proof, we can also conclude that this is true for negative integers  $p$  as well. Finally, by taking into account that the zero tuple  $0 \stackrel{\text{def}}{=} (0, \dots, 0)$  can be represented as  $Z + (-Z)$ , we conclude that  $0 \cdot Z = 0$  also belongs to the set  $S_{\sim}$ .

Thus, if a tuple  $Z$  belongs to the set  $S_{\sim}$ , then for every integer  $p$ , the tuple  $p \cdot Z$  also belongs to the set  $S_{\sim}$ .

9°. Similarly, from Parts 4 and 5 of this proof, it follows that

- if  $Z = (Z_1, \dots, Z_n) \in S_{\succ}$ , then for every positive integer  $p$ , the tuple  $p \cdot Z$  also belongs to the set  $S_{\succ}$ , and
- if  $Z = (Z_1, \dots, Z_n) \in S_{\prec}$ , then for every positive integer  $p$ , the tuple  $p \cdot Z$  also belongs to the set  $S_{\prec}$ .

10°. Let us prove that for every rational number  $r = \frac{p}{q}$ , where  $p$  is an integer and  $q$  is a positive integer, if a tuple  $Z$  belongs to the set  $S_{\sim}$ , then the tuple  $r \cdot Z$  also belongs to the set  $S_{\sim}$ .

Indeed, according to Part 8,  $Z \in S_{\sim}$  implies that  $p \cdot Z \in S_{\sim}$ .

According to Part 2, for the tuple  $r \cdot Z$ , we have either  $r \cdot Z \in S_{\sim}$ , or  $r \cdot Z \in S_{\succ}$ , or  $r \cdot Z \in S_{\prec}$ .

- If  $r \cdot Z \in S_{\succ}$ , then, by Part 9, we would get  $p \cdot Z = q \cdot (r \cdot Z) \in S_{\succ}$ , which contradicts our result that  $p \cdot Z \in S_{\sim}$ .
- Similarly, if  $r \cdot Z \in S_{\prec}$ , then, by Part 9, we would get  $p \cdot Z = q \cdot (r \cdot Z) \in S_{\prec}$ , which contradicts our result that  $p \cdot Z \in S_{\sim}$ .

Thus, the only remaining option is  $r \cdot Z \in S_{\sim}$ . The statement is proven.

11°. Let us now use continuity to prove that for every real number  $x$ , if a tuple  $Z$  belongs to the set  $S_{\sim}$ , then the tuple  $x \cdot Z$  also belongs to the set  $S_{\sim}$ .

Indeed, a real number  $x$  can be represented as a limit of rational numbers:  $r^{(k)} \rightarrow x$ . According to Part 10, for every  $k$ , we have  $r^{(k)} \cdot Z \in S_{\sim}$ , i.e., the tuple

$$Z^{(k)} \stackrel{\text{def}}{=} (\exp(r^{(k)} \cdot Z_1), \dots, \exp(r^{(k)} \cdot Z_n)) \sim (1, \dots, 1). \quad (36)$$

In particular, this means that  $Z^{(k)} \succeq (1, \dots, 1)$ . In the limit,

$$Z^{(k)} \rightarrow (\exp(x \cdot Z_1), \dots, \exp(x \cdot Z_n)) \succeq (1, \dots, 1). \quad (37)$$

By definition of the sets  $S_{\sim}$  and  $S_{\succ}$ , this means that  $x \cdot Z \in S_{\sim}$  or  $x \cdot Z \in S_{\succ}$ .

Similarly, for  $-(x \cdot Z) = (-x) \cdot Z$ , we conclude that  $-x \cdot Z \in S_{\sim}$  or

$$(-x) \cdot Z \in S_{\succ}. \quad (38)$$

If we had  $x \cdot Z \in S_{\succ}$ , then by Part 7 we would get  $(-x) \cdot Z \in S_{\prec}$ , a contradiction. Thus, the case  $x \cdot Z \in S_{\succ}$  is impossible, and we have  $x \cdot Z \in S_{\sim}$ . The statement is proven.

12°. According to Parts 3 and 11, the set  $S_{\sim}$  is closed under addition and under multiplication by an arbitrary real number. Thus, if tuples  $Z, \dots, Z'$  belong to the set  $S_{\sim}$ , their arbitrary linear combination  $x \cdot Z + \dots + x' \cdot Z'$  also belongs to the set  $S_{\sim}$ . So, the set  $S_{\sim}$  is a linear subspace of the  $n$ -dimensional space of all the tuples.

13°. The subspace  $S_{\sim}$  cannot coincide with the entire  $n$ -dimensional space, because then the pre-ordering relation would be trivial. Thus, the dimension of this subspace must be less than or equal to  $n - 1$ . Let us show that the dimension of this subspace is  $n - 1$ .

Indeed, let us assume that the dimension is smaller than  $n - 1$ . Since the pre-ordering is non-trivial, there exist tuples  $y = (y_1, \dots, y_n)$  and  $y' = (y'_1, \dots, y'_n)$  for which  $y \succ y'$  and thus,  $Z = (Z_1, \dots, Z_n) \in S_{\succ}$ , where  $Z_i = \ln \left( \frac{y_i}{y'_i} \right)$ . From  $Z \in S_{\succ}$ , we conclude that  $-Z \in S_{\prec}$ .

Since the linear space  $S_{\sim}$  is a less than  $(n - 1)$ -dimensional subspace of an  $n$ -dimensional linear space, there is a path connecting  $Z \in S_{\succ}$  and  $-Z \in S_{\prec}$  which avoids  $S_{\sim}$ . In mathematical terms, this path is a continuous mapping  $\gamma : [0, 1] \rightarrow R^n$  for which  $\gamma(0) = Z$  and  $\gamma(1) = -Z$ . Since this path avoids  $S_{\sim}$ , every point  $\gamma(t)$  on this path belongs either to  $S_{\succ}$  or to  $S_{\prec}$ .

Let  $\bar{t}$  denote the supremum (least upper bound) of the set of all the values  $t$  for which  $\gamma(t) \in S_{\succ}$ . By definition of the supremum, there exists a sequence  $t^{(k)} \rightarrow \bar{t}$  for which  $\gamma(t^{(k)}) \in S_{\succ}$ . Similarly to Part 11, we can use continuity to prove that in the limit,  $\gamma(\bar{t}) \in S_{\succ}$  or  $\gamma(\bar{t}) \in S_{\sim}$ . Since the path avoids the set  $S_{\sim}$ , we thus get  $\gamma(\bar{t}) \in S_{\succ}$ .

Similarly, since  $\gamma(1) \notin S_{>}$ , there exists a sequence  $t^{(k)} \downarrow \bar{t}$  for which  $\gamma(t^{(k)}) \in S_{<}$ . We can therefore conclude that in the limit,  $\gamma(\bar{t}) \in S_{>}$  or  $\gamma(\bar{t}) \in S_{\sim}$  – a contradiction with our previous conclusion that  $\gamma(\bar{t}) \in S_{>}$ .

This contradiction shows that the linear space  $S_{\sim}$  cannot have dimension smaller than  $n - 1$  and thus, that this space have dimension  $n - 1$ .

14°. Every  $(n - 1)$ -dimensional linear subspace of an  $n$ -dimensional superspace separates the superspace into two half-spaces. Let us show that one of these half-spaces is  $S_{>}$  and the other is  $S_{<}$ .

Indeed, if one of the subspaces contains two tuples  $Z$  and  $Z'$  for which  $Z \in S_{>}$  and  $Z' \in S_{<}$ , then the line segment  $\gamma(t) = t \cdot Z + (1 - t) \cdot Z'$  containing these two points also belongs to the same subspace, i.e., avoids the set  $S_{\sim}$ . Thus, similarly to Part 13, we would get a contradiction.

So, if one point from a half-space belongs to  $S_{>}$ , all other points from this subspace also belong to the set  $S_{>}$ . Similarly, if one point from a half-space belongs to  $S_{<}$ , all other points from this subspace also belong to the set  $S_{<}$ .

15°. Every  $(n - 1)$ -dimensional linear subspace of an  $n$ -dimensional space has the form

$$\alpha_1 \cdot Z_1 + \dots + \alpha_n \cdot Z_n = 0 \quad (39)$$

for some real values  $\alpha_i$ , and the corresponding half-spaces have the form

$$\alpha_1 \cdot Z_1 + \dots + \alpha_n \cdot Z_n > 0 \quad (40)$$

and

$$\alpha_1 \cdot Z_1 + \dots + \alpha_n \cdot Z_n < 0. \quad (41)$$

The set  $S_{>}$  coincides with one of these subspaces. If it coincides with the set of all tuples  $Z$  for which  $\alpha_1 \cdot Z_1 + \dots + \alpha_n \cdot Z_n < 0$ , then we can rewrite it as

$$(-\alpha_1) \cdot Z_1 + \dots + (-\alpha_n) \cdot Z_n > 0, \quad (42)$$

i.e., as  $\alpha'_1 \cdot Z_1 + \dots + \alpha'_n \cdot Z_n > 0$  for  $\alpha'_i = -\alpha_i$ .

Thus, without losing generality, we can conclude that the set  $S_{>}$  coincides with the set of all the tuples  $Z$  for which  $\alpha_1 \cdot Z_1 + \dots + \alpha_n \cdot Z_n > 0$ . We have mentioned that

$$y' = (y'_1, \dots, y'_n) \succ y = (y_1, \dots, y_n) \Leftrightarrow (Z_1, \dots, Z_n) \in S_{>}, \quad (43)$$

where  $Z_i = \ln\left(\frac{y'_i}{y_i}\right)$ . So,

$$y' \succ y \Leftrightarrow \alpha_1 \cdot Z_1 + \dots + \alpha_n \cdot Z_n = \alpha_1 \cdot \ln\left(\frac{y'_1}{y_1}\right) + \dots + \alpha_n \cdot \ln\left(\frac{y'_n}{y_n}\right) > 0. \quad (44)$$

Since  $\ln\left(\frac{y'_i}{y_i}\right) = \ln(y'_i) - \ln(y_i)$ , the last inequality in the formula (44) is equivalent to

$$\alpha_1 \cdot \ln(y'_1) + \dots + \alpha_n \cdot \ln(y'_n) > \alpha_1 \cdot \ln(y_1) + \dots + \alpha_n \cdot \ln(y_n). \quad (45)$$



Let us take exp of both sides of the formula (45); then, due to the monotonicity of the exponential function, we get an equivalent inequality

$$\exp(\alpha_1 \cdot \ln(y'_1) + \dots + \alpha_n \cdot \ln(y'_n)) > \exp(\alpha_1 \cdot \ln(y_1) + \dots + \alpha_n \cdot \ln(y_n)). \quad (46)$$

Here,

$$\exp(\alpha_1 \cdot \ln(y'_1) + \dots + \alpha_n \cdot \ln(y'_n)) = \exp(\alpha_1 \cdot \ln(y'_1)) \cdot \dots \cdot \exp(\alpha_n \cdot \ln(y'_n)),$$

where for every  $i$ ,  $e^{\alpha_i \cdot z_i} = (e^{z_i})^{\alpha_i}$ , with  $z_i \stackrel{\text{def}}{=} \ln(y'_i)$ , implies that

$$\exp(\alpha_i \cdot \ln(y'_i)) = (\exp(\ln(y'_i)))^{\alpha_i} = (y'_i)^{\alpha_i}, \quad (47)$$

so

$$\exp(\alpha_1 \cdot \ln(y'_1) + \dots + \alpha_n \cdot \ln(y'_n)) = (y'_1)^{\alpha_1} \cdot \dots \cdot (y'_n)^{\alpha_n} \quad (48)$$

and similarly,

$$\exp(\alpha_1 \cdot \ln(y_1) + \dots + \alpha_n \cdot \ln(y_n)) = y_1^{\alpha_1} \cdot \dots \cdot y_n^{\alpha_n}. \quad (49)$$

Thus, due to (44), (45), (46), (48), and (49), the condition  $y' \succ y$  is equivalent to:

$$\prod_{i=1}^n y_i^{\alpha_i} > \prod_{i=1}^n (y'_i)^{\alpha_i}. \quad (50)$$

Similarly, we prove that

$$(y_1, \dots, y_n) \sim y' = (y'_1, \dots, y'_n) \Leftrightarrow \prod_{i=1}^n y_i^{\alpha_i} = \prod_{i=1}^n (y'_i)^{\alpha_i}. \quad (51)$$

The condition  $\alpha_i > 0$  follows from our assumption that the pre-ordering is monotonic.

The theorem is proven.